ASSET CREATOR(s): Johanna Loomba (UVA), Saad Ljazouli (Palantir), Andrea Zhou (UVA), Suchetha Sharma (UVA), Ethan Asseffa (UVA), Evan French (VCU) with ideation and concept set contributions from the Logic Liaisons, Data Liaisons, and N3C Domain Team leads.

README SECTIONS:

TEMPLATE STATUS:

AS OF JAN, 2022: The underlying concept sets used for conditions in this template were built using key concepts in SNOMED-CT plus their descendants. As per direction from the N3C PIs, we plan to transition to externally validated concept sets that are being curated by the data liaisons. That change will be noted here once made.

AS OF MAR, 2022: The template was modified to include COVID-19 diagnosed patients in addition to COVID-19 PCR or AG lab positive patients. See the release notes section in this document for additional details.

AS OF APR, 2022: The template has been updated to accept input from a fusion sheet/manual entry dataset to allow researchers to more easily adapt the template to read in their desired concept sets. See the release notes section in this document for additional details.

PURPOSE & USE:

This set of templates (Level 2 and Level 3 data versions) provides sample code and summary datasets including a day-level and a patient-level table. The day-level "all facts" table has a single row for each patient and each day along with flags for visit relationship to COVID-19 index event or COVID-19 related events. The patient-level table contains one row for each PCR or AG lab positive COVID-19 patient or U07.1 diagnosed COVID-19 patient with a number of commonly referenced facts and indicators derived from the N3C datasets. The template outputs can be used in a variety of ways:

1) **ANALYZE IN CONTOUR:** You can analyze the output datasets directly in Palantir's Contour App. Simply open the LDS (Level 3 projects) or the De-ID (Level 2 projects) folder (links located just underneath where the link for this READ ME was located), select the dataset you want to use (person-level or day-level "all facts") then select "Analyze" in the upper-right to open the corresponding dataset in Contour. **The datasets in the knowledge object are updated with each refresh of the OMOP data in N3C.**

2) **IMPORT TEMPLATE OUTPUTS:** You can import the day-level and/or patient-level datasets that are produced by these templates directly into your code workbook or a new code workbook. **The datasets in the knowledge object are updated with each refresh of the OMOP data in N3C.**

To import into your existing code workbook, simply open the LDS (L3 projects) or De-ID (L2 projects) dataset to copy the file path next to "Location" in the middle left properties panel using the clipboard symbol that appears when hovering over the path. Then, open your code workbook, select "Import dataset" in the upper-left, and paste the file path into the search bar.

To analyze in a new code workbook, go to the "Actions" menu in upper-right after opening the desired dataset and select "Analyze in Code Workbook". Once in the correct workbook, the dataset can be used directly for analysis or joined with other tables you have created (for example, a cohort table you have created under your DUR).

3) **IMPORT AND CUSTOMIZE TEMPLATE:** Alternatively, you can import the appropriate code template and modify the code so that the resulting tables reflect the

specific cohort and derived facts that your DUR requires for its analyses. Detailed instructions are provided below for this process.

NOTES ABOUT VARIABLE CREATION

SDOH (Social Determinants of Health) DATA:

SDOH variables can be joined based on person_id using the output dataset of this template which provides variables at the ZCTA and County level as recommended by the SDOH domain team.

COMORBIDITY indicators before and post- COVID:

The _post_COVID_indicator columns have a value of 1 if COVID-19 patient has any record of given comorbidity on any day after the index event. The _before_or_day_of_COVID_indicator columns have a value of 1 if COVID-19 patient has any record of given comorbidity **on the same day or prior to the index event**. This decision was made because, for patients whose medical history was captured for the first time on their COVID index date (no prior visits recorded), most chronic conditions captured that day pre-dated infection. HOWEVER, the user should consider excluding conditions flagged as “before or day of COVID” if the variables you choose to generate are less chronic, such as fever. To do this, exclude “before or day of COVID” conditions when the **observation_period_before_COVID** is zero. To utilize the facts found on the index date as part of the "during COVID" or "post COVID" periods, import this template to allow for study-specific fact indexing of the COVID index or another index date of interest.

DEFAULT Concept Sets Used:

This set of comorbidities includes conditions listed in the Charlson Comorbidity Index. However, instead of using Quan et. al list of ICD codes only, we instead created these concept sets using the primary conditions listed in SNOMED-CT hierarchy and including all descendants.

We also created concept sets for each of the comorbidity sections listed on this CDC website as of Dec 2021: CDC "People of any age with the conditions listed below are more likely to get severely ill from COVID-19". Notes regarding each concept set and

validation can be found in the Concept Set Browser. Search the comorbidity as named in the variable, but include spaces between words. For example, to look up the concept set used for the "HEARTFAILURE_before_or_day_of_covid_indicator", search "HEART FAILURE" in the concept set browser. Johanna Loomba is the creator of the concept sets used, so you can also use this for your search.

DATA DICTIONARY (for variables in final patient-level table):

All concept sets in the below descriptions are marked by italics. You can swap out concept names in the code template's fusion sheet/manual entry dataset to use your preferred source concept sets for any of the below variables.

person_id, string: Unique identifier for the N3C patient

COVID_first_PCR_or_AG_lab_positive, date: Date representing first instance of measurement "*ATLAS SARS-CoV-2 rt-PCR and AG*" with a positive result "*ResultPos*"

COVID_first_diagnosis_date, date: Date representing the first instance of a COVID-19 diagnosis *"N3C Covid Diagnosis"* having been charted (when available)

COVID_first_poslab_or_diagnosis_date, date: Date representing the first instance of either the measurement "*ATLAS SARS-CoV-2 rt-PCR and AG*" with a positive result "*ResultPos*" or a COVID-19 diagnosis *"N3C Covid Diagnosis"* having been charted (when available). THIS DATE SERVES AS THE INDEX DATE FOR MOST OF THE BELOW VARIABLES.

number_of_visits_before_covid, long: The total number of visit days the patient was recorded having prior to their COVID index event visit, with hospitalizations collapsed to a single visit

observation_period_before_covid, integer: The number of days between the patient's earliest recorded visit and the index event

number_of_visits_post_covid, long: The total number of visit days the patient was recorded having after their COVID index event visit, with hospitalizations collapsed to a single visit

observation_period_post_covid, integer: The number of days between the patient's index event and last recorded visit

sex, string: gender_concept_name from main OMOP Person Table. Per OMOP ETL conventions defined here, "use the gender or sex value present in the data under the assumption that it is the biological sex at birth. This field should not be used to study gender identity issues. If the source data captures gender identity it should be stored in the OBSERVATION table." Per OHDSI documentation here, "the Gender domain captures all concepts about the sex of a person, denoting the biological and physiological characteristics."

city, string: City from main OMOP location table

state, string: State from main OMOP location table

postal_code, string: Zip from main OMOP location table

county, string: County from main OMOP location table

age_at_covid, long: Calculated age using the year of birth (for L2) or date of birth (for L3) and the date of the index event. As of 7/15/22, July 1st used as placeholder month and day of birth when there are 0s or nulls in the OMOP person table to avoid biasing towards older age.

race, string: "Hispanic or Latino" for anyone with this race, otherwise "Asian", "Black or African American", “Native Hawaiian or Other Pacific Islander”, "American Indian or Alaska Native", "White", or "Other". "Unknown" is assigned to all others. **NOTE**: A number of patients with a source value of Asian or Other Pacific Islander are categorized as “Unknown” until the site is able to differentiate and correct their source values since the current value does not reasonably fall into only one of the 6 minimum race/ethnicity categories defined by the NIH.

race_ethnicity, string: "Hispanic or Latino Any Race" for anyone with this ethnicity, otherwise "Asian Non-Hispanic", "Black or African American Non-Hispanic", “Native Hawaiian or Other Pacific Islander Non-Hispanic”, "American Indian or Alaska Native Non-Hispanic", "White Non-Hispanic", or "Other Non-Hispanic". "Unknown" is assigned to all others. **NOTE**: A number of patients with a source value of Asian or Other Pacific Islander are categorized as “Unknown” until the site is able to differentiate and correct their source values since the current value does not reasonably fall into only one of the 6 minimum race/ethnicity categories defined by the NIH.

data_partner_id, integer: The anonymized institution contributing this patient's record to N3C

data_extraction_date, date: The date that the patient's site last ran data extraction scripts for N3C

cdm_name, string: The Clinical Data Model (CDM) associated with this patient's contributing institution (OMOP, TriNetX, ACT, or PCORNet). The different CDMs have different strengths and weaknesses and may result in a pattern of missingness.

cdm_version, string: The version of the Clinical Data Model used by this patient's contributing institution

shift_date_yn, string: Y if the patient's site shifts dates, otherwise N (note that for De-Id L2 data, all the records are shifted, so this field has been uniformly changed to reflect Y for all sites)

max_num_shift_days, string: The maximum number of days that the patient's site shifts dates, otherwise N (note that for De-Id L2 data, all records are shifted, so this field has been uniformly changed to reflect +180 days for all sites)

BMI_max_observed_or_calculated_before_or_day_of_covid, double: The maximum Body Mass Index on the same day or prior to the index event. Default parameters of reasonability are set for height, weight, and BMI, but can be controlled by template user as input parameters. The max reasonable BMI measure "*body mass index"* prior to the index date is reported. Both calculated BMI (using a weight "*Body weight (LG34372-9 and SNOMED)"* and temporally associated height "*Height (LG34373-7 + SNOMED)"* for calculation) and reported BMI as a measure are used to identify the closest measure. If both are available on the same date, the EMR reported BMI is used instead of calculated BMI.

OBESITY_before_or_day_of_covid_indicator, integer: Value of 1 if COVID-19 patient has any record of "OBESITY" before or day of index event OR has BMI_max_observed_or_calculated_before_or_day_of_covid > 30.

SYSTEMICCORTICOSTEROIDS_before_or_day_of_covid_indicator, integer: Value of 1 if COVID-19 patient has any record (order or administration) of *“N3C CORTICOSTEROIDS FOR SYSTEMIC USE”* drug prior to the index event. See the "Logic Liaison Tips" section for more information about medications.

Antibody_Neg_before_or_day_of_covid_indicator, integer: Value of 1 if COVID-19 patient has any record of *"Atlas #818 [N3C] CovidAntibody retry"* with *"ResultNeg"* value on the same day or prior to the index event.

Antibody_Pos_before_or_day_of_covid_indicator, integer: Value of 1 if COVID-19 patient has any record of *"Atlas #818 [N3C] CovidAntibody retry"* with *"ResultPos"* value on the same day or prior to the index event.

number_of_COVID_vaccine_doses_before_or_day_of_covid, long: Count of COVID-19 vaccine doses received by COVID-19 patient on the same day or prior to the index event. See community note regarding N3C vaccination data considerations.

COVID_diagnosis_during_covid_hospitalization_indicator, integer: Value of 1 if the patient had a COVID-19 diagnosis charted between first_COVID_hospitalization_start_date and first_COVID_hospitalization_end_date

LL_IMV_during_covid_hospitalization_indicator, integer: Value of 1 if patient had observation, procedure, device_exposure, or condition related to invasive mechanical ventilation *"[ICU/MODS]IMV"* between first_COVID_hospitalization_start_date and first_COVID_hospitalization_end_date

LL_ECMO_during_covid_hospitalization_indicator, integer: Value of 1 if patient had observation, procedure, device_exposure, or condition related to *"Kostka - ECMO"* between first_COVID_hospitalization_start_date and first_COVID_hospitalization_end_date

COVIDREGIMENCORTICOSTEROIDS_during_covid_hospitalization_indicator, integer: Value of 1 if COVID-19 patient has any record (order or administration) of "*NIH Systemic Corticosteroids"* drug between first_COVID_hospitalization_start_date and first_COVID_hospitalization_end_date. This concept set was created to reflect Systemic Corticosteroids recommended by NIH for COVID treatment, only oral and IV . NIH Oct 19, 2021 list. **NOTE:** These medications are also often used to treat diseases other than COVID-19. See the "Logic Liaison Tips" section for more information about medications.

REMDISIVIR_during_covid_hospitalization_indicator, integer: Value of 1 if COVID-19 patient has any record (order or administration) of "*Remdesivir"* drug between first_COVID_hospitalization_start_date and first_COVID_hospitalization_end_date. See the "Logic Liaison Tips" section for more information about medications.

COVID_patient_death_during_covid_hospitalization_indicator, integer: Value of 1 if patient had death recorded between first_COVID_hospitalization_start_date and first_COVID_hospitalization_end_date

BMI_max_observed_or_calculated_post_covid, double: The maximum Body Mass Index after the index event. Default parameters of reasonability are set for height, weight, and BMI, but can be controlled by template user as input parameters. The max reasonable BMI measure "*body mass index"* after the index date is reported. Both calculated BMI (using a weight *"Body weight (LG34372-9 and SNOMED)"* and temporally associated height *"Height (LG34373-7 + SNOMED)"* for calculation) and

reported BMI as a measure are used to identify the closest measure. If both are available on the same date, the EMR reported BMI is used instead of calculated BMI.

OBESITY_post_covid_indicator, integer: Value of 1 if COVID-19 patient has any record of "OBESITY" post index event OR has BMI_max_observed_or_calculated_post_covid > 30.

Long_COVID_diagnosis_post_covid_indicator, integer: Value of 1 if the patient has Long COVID Diagnosis U09.9 (concept set *"Long-COVID (PASC)"*)

Long_COVID_clinic_visit_post_covid_indicator, integer: Value of 1 if the patient had any known *"Long COVID Clinic Visit"* [note that not all sites are transmitting this data]

MISC_post_covid_indicator, integer: Value of 1 if COVID-19 patient has any record of *"MULTISYSTEM INFLAMMATORY SYNDROME - CHILDREN"* after the index event.

PNEUMONIADUETOCOVID_post_covid_indicator, integer: Value of 1 if COVID-19 patient has any record of *"PNEUMONIA DUE TO COVID-19"* after the index event.

SYSTEMICCORTICOSTEROIDS_post_covid_indicator, integer: Value of 1 if COVID-19 patient has any record (order or administration) of *"N3C CORTICOSTEROIDS FOR SYSTEMIC USE"* drug following the index event. See the "Logic Liaison Tips" section for more information about medications.

Antibody_Neg_post_covid_indicator, integer: Value of 1 if COVID-19 patient has any record of *"Atlas #818 [N3C] CovidAntibody retry"* with *"ResultNeg"* value after the index event.

Antibody_Pos_post_covid_indicator, integer: Value of 1 if COVID-19 patient has any record of *"Atlas #818 [N3C] CovidAntibody retry"* with *"ResultPos"* value after the index event.

PCR_AG_Neg_post_covid_indicator, integer: Value of 1 if COVID-19 patient has any record of *"ATLAS SARS-CoV-2 rt-PCR and AG"* with *"ResultNeg"* value after the index event.

PCR_AG_Pos_post_covid_indicator, integer: Value of 1 if COVID-19 patient has any record of *"ATLAS SARS-CoV-2 rt-PCR and AG"* with *"ResultPos"* value after the index event.

number_of_COVID_vaccine_doses_post_covid, long: Count of COVID-19 vaccine doses received by COVID-19 patient after the index event. See community note regarding N3C vaccination data considerations.

had_at_least_one_reinfection_post_covid_indicator, integer: Value of 1 if COVID-19 patient had a positive PCR or AG test > 60 days after the earliest_covid_lab_or_diagnosis date

first_COVID_ED_only_start_date, date: Date of first qualifying Emergency Department visit (see COVID_associated_ED_only_visit_indicator)

first_COVID_hospitalization_start_date, date: Start date of first qualifying hospitalization (see COVID_associated_hospitalization_indicator)

first_COVID_hospitalization_end_date, date: End date of first qualifying hospitalization (see COVID_associated_hospitalization_indicator)

COVID_hospitalization_length_of_stay, integer: The number of days from first_COVID_hospitalization_start_date to first_COVID_hospitalization_end_date

COVID_associated_ED_only_visit_indicator, integer: Value of 1 if the patient had an Emergency Department visit *"[PASC] ED Visits"* AND a COVID-19 diagnosis *"N3C Covid Diagnosis"* charted in the day prior to 16 days following the index event. This time frame is aligned with the windows used by the CDC. It can be modified in the template. Note also that ED visits that overlap with an admission are considered part of the hospitalization (see microvisit_to_macrovisit table documentation in the enclave), so those patients who first present in the ED but are then admitted instead of discharged will NOT be flagged here. This is also aligned with CDC methods for counting the COVID-19 associated ED visits.

COVID_associated_hospitalization_indicator, integer: Value of 1 if the patient was hospitalized in the day prior to 16 days following the index event AND a COVID-19 diagnosis *"N3C Covid Diagnosis"* charted in the day prior to 16 days following the index event. This time frame is aligned with the windows used by the CDC. It can be modified in the template. Note that the N3C microvisit_to_macrovisit table was used to identify hospitalizations (see enclave documentation here).

COVID_patient_death_indicator, integer: Value of 1 if the patient had death or discharge to hospice recorded in N3C.

death_within_specified_window_post_covid, integer: Value of 1 if the patient had death date occur within specified window following COVID_first_PCR_or_AG_lab_positive. Specified window is defaulted to 60 days which is also utilized for specifying the window in which to look for a reinfection following index event. This can be defined by the template user via the reinfection_wait_time parameter in the cohort_all_facts_table node.

Severity_Type, string: See Community Note "Logic Liaison: Severity Type Definition"

COMORBIDITY indicators before or day of and post- COVID:

The _before_or_day_of_COVID_indicator columns have a value of 1 if COVID-19 patient has any record of given comorbidity on the same day or prior to the index event. The _post_COVID_indicator columns have a value of 1 if COVID-19 patient has any record of given comorbidity on any day after the index event.

TUBERCULOSIS_before_or_day_of_covid_indicator / TUBERCULOSIS_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"TUBERCULOSIS"* before or day of / post index event

MILDLIVERDISEASE_before_or_day_of_covid_indicator / MILDLIVERDISEASE_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"MILD LIVER DISEASE*" before or day of / post index event

MODERATESEVERELIVERDISEASE_before_or_day_of_covid_indicator / MODERATESEVERELIVERDISEASE_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"MODERATE OR SEVERE LIVER DISEASE*" before or day of / post index event

THALASSEMIA_before_or_day_of_covid_indicator / THALASSEMIA_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"THALASSEMIA"* before or day of / post index event

RHEUMATOLOGICDISEASE_before_or_day_of_covid_indicator / RHEUMATOLOGICDISEASE_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"RHEUMATOLOGIC DISEASE"* before or day of / post index event

DEMENTIA_before_or_day_of_covid_indicator / DEMENTIA_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"DEMENTIA"* before or day of / post index event

CONGESTIVEHEARTFAILURE_before_or_day_of_covid_indicator / CONGESTIVEHEARTFAILURE_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"CONGESTIVE HEART FAILURE*" before or day of / post index event

SUBSTANCEUSEDISORDER_before_or_day_of_covid_indicator / SUBSTANCEUSEDISORDER_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"SUBSTANCE USE DISORDER"* before or day of / post index event

DOWNSYNDROME_before_or_day_of_covid_indicator / DOWNSYNDROME_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"DOWN SYNDROME"* before or day of / post index event

KIDNEYDISEASE_before_or_day_of_covid_indicator / KIDNEYDISEASE_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"KIDNEY DISEASE"* before or day of / post index event.

MALIGNANTCANCER_before_or_day_of_covid_indicator / MALIGNANTCANCER_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"MALIGNANT CANCER"* before or day of / post index event

DIABETESCOMPLICATED_before_or_day_of_covid_indicator / DIABETESCOMPLICATED_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"DIABETES COMPLICATED"* before or day of / post index event

CEREBROVASCULARDISEASE_before_or_day_of_covid_indicator / CEREBROVASCULARDISEASE_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"CEREBROVASCULAR DISEASE"* before or day of / post index event

PERIPHERALVASCULARDISEASE_before_or_day_of_covid_indicator / PERIPHERALVASCULARDISEASE_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"PERIPHERAL VASCULAR DISEASE"* before or day of / post index event

PREGNANCY_before_or_day_of_covid_indicator / PREGNANCY_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"PREGNANT*" before or day of / post index event

HEARTFAILURE_before_or_day_of_covid_indicator / HEARTFAILURE_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"HEART FAILURE"* before or day of / post index event

HEMIPLEGIAORPARAPLEGIA_before_or_day_of_covid_indicator / HEMIPLEGIAORPARAPLEGIA_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"HEMIPLEGIA or PARAPLEGIA"* before or day of / post index event.

PSYCHOSIS_before_or_day_of_covid_indicator / PSYCHOSIS_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"PSYCHOSIS"* before or day of / post index event

CORONARYARTERYDISEASE_before_or_day_of_covid_indicator / CORONARYARTERYDISEASE_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"CORONARY ARTERY DISEASE"* before or day of / post index event

DEPRESSION_before_or_day_of_covid_indicator / DEPRESSION_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"DEPRESSION"* before or day of / post index event

METASTATICSOLIDTUMORCANCERS_before_or_day_of_covid_indicator / METASTATICSOLIDTUMORCANCERS_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"METASTATIC SOLID TUMOR CANCERS"* before or day of / post index event

HIVINFECTION_before_or_day_of_covid_indicator / HIVINFECTION_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"HIV INFECTION"* before or day of / post index event

CHRONICLUNGDISEASE_before_or_day_of_covid_indicator / CHRONICLUNGDISEASE_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"CHRONIC LUNG DISEASE*" before or day of / post index event

PEPTICULCER_before_or_day_of_covid_indicator / PEPTICULCER_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"PEPTIC ULCER"* before or day of / post index event

SICKLECELLDISEASE_before_or_day_of_covid_indicator / SICKLECELLDISEASE_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"SICKLE CELL DISEASE"* before or day of / post index event

MYOCARDIALINFARCTION_before_or_day_of_covid_indicator / MYOCARDIALINFARCTION_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"MYOCARDIAL INFARCTION*" before or day of / post index event

DIABETESUNCOMPLICATED_before_or_day_of_covid_indicator / DIABETESUNCOMPLICATED_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"DIABETES UNCOMPLICATED"* before or day of / post index event

CARDIOMYOPATHIES_before_or_day_of_covid_indicator / CARDIOMYOPATHIES_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"CARDIOMYOPATHIES"* before or day of / post index event

HYPERTENSION_before_or_day_of_covid_indicator / HYPERTENSION_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"HYPERTENSION"* before or day of / post index event

OTHERIMMUNOCOMPROMISED_before_or_day_of_covid_indicator / OTHERIMMUNOCOMPROMISED_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"IMMUNODEFICIENCY*" before or day of / post index event

PULMONARYEMBOLISM_before_or_day_of_covid_indicator / PULMONARYEMBOLISM_post_covid_indicator , integer: Value of 1 if COVID-19 patient has condition *"PULMONARY EMBOLISM"* before or day of / post index event

TOBACCOSMOKER_before_or_day_of_covid_indicator / TOBACCOSMOKER_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"TOBACCO SMOKER"* before or day of / post index event

SOLIDORGANORBLOODSTEMCELLTRANSPLANT_before_or_day_of_covid_indicator / SOLIDORGANORBLOODSTEMCELLTRANSPLANT_post_covid_indicator, integer: Value of 1 if COVID-19 patient has condition *"TRANSPLANT OF SOLID ORGAN OR BLOOD STEM CELL"* before or day of / post index event

TEMPLATE LIMITATIONS

Because our goal is to provide a widely usable set of variables with a well-documented and easy-to-understand approach, we have not customized the underlying concept sets to meet the needs of any specific project. Instructions for modifying the template are provided below should you want to replace the concept sets used with your own versions. Other limitations of the template reflect common limitations of this data, some of which are described in the tips below.

INTRODUCTORY LOGIC LIAISON TIPS

(see N3C Community Note here)

HOW TO MODIFY THIS TEMPLATE FOR A SPECIFIC N3C PROJECT

You may wish to modify input parameters (such as acceptable BMI range, for example), underlying concept sets, or time frames used (say to define hospitalization). You may also want to add additional variables or join the template output tables with a cohort or other data frame you have produced. The below section describes how to import the code template and convert to code transform wherever code customization is required.

Import Template

- Open a new code workbook and select “Skip this Step” under "Import Dataset".
- Make sure you are using the "default" or "high-memory" environment (selected under the environment menu towards the top middle of the window: Customize Spark environment → profile-high-memory on the left panel of pop-up → Update Spark environment).
- Click “New Transform”, select “Templates”, search for the “[LOGIC LIAISON TEMPLATE] L2 and L3 Fact Tables: COVID-19 Diagnosed or Lab Confirmed Patients" and import into the workbook. When you press the ‘apply transformation’ button, a box pops up that says that there are resources within the template that are not within the scope of the project. You can agree to add these resources.
- Once the template nodes appear, select "zoom to fit" icon from lower-left corner so that the transforms all fit on your screen.
- *Y*ou will need to review and edit each transformation as needed in each step below.
- **To just view the code**, select a transformation node and click the “Toggle View” button in upper-right of the control window that appears at the bottom half of the screen
- **To change the code**, select **“**Actions” button to the right of “Toggle View” **i**n the transformation node, and click **"Convert to code transform"**. **Note: once converted to code transform, you will not be able to update your imported template with the most recent version of the Knowledge Store template while keeping your changes made within the same nodes.*

Import OMOP Data

- The template selected will import the domain tables for the level of data chosen. The L2 template will automatically import deidentified domain tables and the L3 template will automatically import the limited dataset domain tables. These tables will appear on the left side. Similarly, we use the microvisit_to_macrovisit table for identifying hospitalizations (see enclave documentation here) as opposed to the visit_occurrence table.

Import Fusion Sheet/Manual Entry Dataset

- The template will default to importing the dataset generated by the "SNOMED" tab of a fusion sheet. This set of concept sets are ones developed using a small selection of parent SNOMED concepts and including all SNOMED-CT descendants.
- To customize this list of concept sets used in the fact table, you should copy and paste the default LL_concept_sets_fusion dataset into a fusion sheet or manual entry dataset to edit. At this time, you can only add/remove concepts for the following OMOP domains: conditions, procedures, devices, observations, and medications. **It is required to keep IMV and ECMO in the modified version of this list.** Ensure you specify all four necessary fields using the appropriate information:
- **concept_set_name** should be exactly as it is found in the concept set browser. The code will default to the most recent version of the specified concept set. This can be modified to use codeset ids with appropriate corresponding code changes in the nodes.
- **indicator_prefix** field is whatever you wish to name the variable
- **domain** is the OMOP table(s)/domain(s) in which the concepts of the concept set specified in column A can be found
- **pre_during_post** will determine the timeframe around the COVID index date in which the fact will be searched for as well as the indicator suffix (_before_or_day_of_COVID_indicator , during_covid_associated_hospitalization_indicator, or _post_covid_indicator) that will be appended to indicator_prefix when generating the patient summary table.
- Once you have a new input table with your chosen concept sets, connect your new table as input to the **customize_concept_sets** node by clicking the white box with "LL_concept_sets_fusion_SNOMED" displayed in the template view of the node. Once the box has turned a light orange, select your new table to set it as the input (no manual changes needed to the code!).

- Running the pipeline will now create event flags to the visit-level table for all the concept sets you defined in the customized input and will generate pre-COVID, during COVID hospitalization, and post-COVID flags in the summary table for these same concepts.
- If you wish to keep all of the ones created by the template as is, it is recommended to import the desired visit-level or patient-level datasets from the Knowledge Store and only run the template pipeline for your customized concept sets. When using this method, it is **still required to keep IMV and ECMO in this custom list.**

COHORT Node

- This node identifies all patients with positive results from a PCR or AG COVID-19 lab test and the date of the patients' first instance of this type of COVID-19+ test. It also identifies all patients with a COVID-19 diagnosis charted and the date of the patients' first instance of this type of diagnosis (when available). The earlier of the two is considered the index date for downstream calculations.
- This transform also gathers some commonly used facts about these patients from the "person" and "location" tables, as well as some facts about the patient's institution (from the "manifest" table). Available age, race, ethnicity, and location data is gathered at this node. The patient's number of visits before and after covid as well as the number of days in their observation period before and after covid is calculated from the "microvisits_to_macrovisits" table in this node. These facts are joined with the final patient-level table in the final node.
- **IMPORTANT NOTE FOR ANALYSTS:** If you are modifying code or adding to this pipeline, you might consider setting the value of the 'proportion_of_patients_to_use' variable to something <1.0. This allows you to work with a small sample of the patients while testing. You can change this value back to 1.0 once done debugging your code.

Find Facts Related to the Patients in Cohort

Note that the below nodes all filter rows to those patients that are identified in the COHORT node.

- The **conditions_of_interest** node filters the condition_occurrence table for rows that have a condition_concept_id associated with one of the concept sets described in the above data dictionary through the use of a fusion sheet/manual entry dataset. Indicator names for these conditions are

assigned, and the indicators are collapsed to unique instances on the basis of patient and date.

- The **observations_of_interest, drugs_of_interest, procedures_of_interest, and devices_of_interest** nodes filter the source OMOP tables for rows that have a standard concept id associated with one of the concept sets described in the data dictionary through the use of a fusion sheet/manual entry dataset. Indicator names for these variables are assigned, and the indicators are collapsed to unique instances on the basis of patient and date.
- The **measurements_of_interest** node filters the measurements table for rows that have a measurement_concept_id associated with one of the concept sets described in the above data dictionary. It finds the harmonized value as a number for the quantitative measurements and collapses these values to unique instances on the basis of patient and date. It also finds the value as concept id for the qualitative measurements (covid labs) and collapses these to unique instances on the basis of patient and date.
- Measurement BMI cutoffs included are intended for adults. Analyses focused on pediatric measurements should use different bounds for BMI measurements. Note that the code looks for both BMI that is directly recorded as well as BMI that is calculated. If both are in a reasonable range and have the same date associated, the recorded BMI is used. Please note that default "reasonable" high low limits are as follows, but can be controlled by the user as a template parameter. Just click on this template node to edit the input parameters and rerun the workbook.
- reasonable BMI: 10 - 100
- reasonable Height: 0.6 - 2.43 meters
- reasonable Weight: 5-300 kilograms
- The **visits_of_interest** node queries the microvisits_to_macrovisits table to identify hospitalizations (see enclave documentation here). The input table can be changed to the visits table if indicated, but the code in the transform would need to be modified accordingly. The parameter called covid_associated_hospitalization_requires_lab_AND_diagnosis is created and allows the user to easily change whether they define COVID-19 associated ED visits and hospitalizations using the CDC definition (lab positive with a COVID-19 diagnosis charted) OR using anyone who is either lab positive or has a COVID-19 diagnosis charted.
- when parameter =True: Per CDC definitions of a COVID-19 associated ED or hospital admission visit, ensure that a COVID-19 diagnosis and ED/hospital admission occurred in the 16 days after or 1 day prior to the PCR or AG positive test (index event).

- when parameter =False: ED or hospital admission visits flagged based on the first instance of a positive COVID-19 PCR or AG lab result OR the first instance of a charted COVID-19 diagnosis when there is no positive lab result within specified timeframe of ED/hospital admission.
- Also, change the time windows around the positive COVID-19 test (OR earliest index event depending on parameter setting) and the macrovisit (hospitalization) start and end date if you wish to define the timeframe differently. The same applies to timelines around ED visits. Number of days between a patient's diagnosis date and their positive lab result is also calculated in this node.
- The **death_persons** node code drops any duplicates from the death table and creates a flag for whether a patient has died.

Assemble Cohort All Facts Table (visit day table)

- All facts collected in the previous steps are combined in the **cohort_all_facts_table** on the basis of unique patient visit days and/or start dates from OMOP tables. Logic is applied to see if the instance of the coded fact appeared in the EHR prior to or after the patient's first COVID-19 positive PCR or AG lab test. Indicators are created for the presence or absence of events, medications, conditions, measurements, device exposures, observations, procedures, and outcomes, either occurring before or day of COVID index date, during the patient's hospitalization, or in the period after COVID index date. To utilize the facts found on the index date as part of the during the patient's hospitalization or in the period after the COVID index date, import this template to allow for study-specific fact indexing of the COVID index or another index date of interest.
- This node also creates an indicator for whether the date where a fact was noted occurred during any hospitalization, not just the COVID associated ones found in the visits of interest node. A flag for the date in which the patient is noted to have their first covid reinfection is also thrown in this node. The default time range is 60 days after index date, but the number of days can be specified by parameter input based on researcher interests. A death within this specified window flag is also thrown in this node and used to calculate severity of this index infection in the summary table.
- This table is useful if the analyst needs to use actual dates of events as it provides more detail than the final patient-level table. Use the max and min functions to find the first and last occurrences of any events.

Create Final Patient Level Table

- The final step is to aggregate information to create a data frame that contains a single row of data for each patient in the cohort.
- The right-most node in the code workbook (**COVID_Patient_Summary_Table**) aggregates all information from the **cohort_all_facts_table** and summarizes each patient's facts in a single row. The patient's hospitalization length of stay is calculated in this node. For patients with ED visits and/or hospitalizations concurrent with their positive COVID-19 index date, indicators are created in this node. This transformation then joins the before or day of COVID, during hospitalization, and post COVID indicator data frames on the basis of unique patients.

UPDATING TO LATEST TEMPLATE VERSION:

- See N3C Community Note titled LL Technical Tip: Updating multi-node template versions

RELEASE NOTES:

TEMPLATE RELEASE NOTES (VERSION CHANGES):

Feb. 2, 2022:

- Added SDOH variables to L3 Template based on recommendations from the Social Determinants of Health Domain Team leadership.
- Modified a handful of variable names in the summary table for case consistency.
- Fixed bugs related to MI, Paralysis, and ECMO variable creation. Changed BMI values of 0 to null.
- Removed optional "Patterns of Missingness by Site". This has been moved into a new Logic Liaison code template under development that will help users assess data quality. Contact Johanna Loomba if you need this code in the meantime.

Mar. 7, 2022:

Modifications:

- before_COVID and post_COVID timeframes changed to pivot around the earlier date between a patient's COVID_PCR_or_AG_lab_positive and COVID_first_diagnosis_date values rather than basing it solely on the COVID_PCR_or_AG_lab_positive date
- Parameterized the time windows around the positive COVID-19 test (OR earliest index event depending on parameter setting) and the macrovisit (hospitalization) start and end date
- Removed flags for COVID_diagnosis_before_covid_indicator and COVID_diagnosis_post_covid_indicator
- Replaced COVID_lab_positive_and_diagnosed_ED_visit_indicator → COVID_associated_ED_only_visit_indicator, COVID_lab_positive_and_diagnosed_hospitalization_indicator → COVID_associated_hospitalization_indicator, and DEATH_during_covid_hospitalization_indicator → COVID_patient_death_during_covid_hospitalization_indicator. Also modified a handful of other variable names slightly to allow for suffix consistency for _before_covid_indicators, _during_covid_hospitalization_indicators, and _post_covid_indicators.
- Switched REMDESIVIR and COVIDREGIMENCORTICOSTEROIDS to be during Covid associated hospitalization indicators rather than post Covid indicators

Bug fixes:

- TOBACCOSMOKER flag was updated to include instances found in the observations and conditions domain rather than just conditions domain alone.

Additions:

- Created new index event date variable that finds the earliest of either the first positive COVID-19 lab result or a COVID-19 diagnosis to use for pre_covid, during_hospitalization, and post_covid timeframes
- Added parameter and corresponding if else statement that allows the user to choose the desired cohort definition for COVID-19 related stand-alone ED visits and hospitalizations. The default code requires a PCR or AG lab positive result and COVID-19 diagnosis in the day prior to 16 days after an ED visit or hospitalization as the index event for inclusion to the cohort (CDC definition). When the parameter is set to False, the code expands this index event requirement to include patients with only a charted COVID-19

diagnosis in the day prior to 16 days after, depending on the earliest possible index event between diagnosis and lab positive.
- Created a flag for MODERATESEVERELIVERDISEASE as an additional condition of interest
- Added two columns to the final patient summary table that counts the number of visits before COVID and post COVID. Hospital stays are collapsed to count as one visit.
- Created a flag in the visit level table for whether observations on a given date occurred during ANY hospitalization stay for a particular patient

Data Cleaning:

- Filtered out unreasonable COVID patient death dates that were prior to 01-01-2018 and after 01-01-2099 for visit level facts table
- Corrected continuous SDOH values filled with 0 to null.
- Changed max_num_shift_days values of NA/na/empty string to 0. Left UNKNOWN as is since that is an ambiguous indicator of date shift.
- FOR L2 only: Added a line of code to add 180 days to max_num_shift_days per site to note systematic shift by N3C

Several code readability and performance optimization changes were also made throughout the template.

Apr. 21, 2022:

Modifications:

- The All Facts table no longer includes any visit date where a fact was recorded in any of the OMOP tables. It now returns only visit dates where any of the facts searched are found to be present.
- Corrected observation period calculations so that it uses only dates from the visit domain which start in 2018 as opposed to across all domains which had numerous abnormal dates resulting in unreasonable observation period values.
- Visit count and observation period calculations have been updated to strictly count visits/days before and after the patient's index date, not including that day of the index event being charted to allow for differentiation between chronic and acute conditions of interest.

Additions

- Template now accepts fusion sheet/manual entry dataset input for concept sets used in conditions_of_interest, observations_of_interest, procedures_of_interest, devices_of_interest, and drugs_of_interest. It also accepts fusion sheet/manual entry dataset input for defining which variables are evaluated in the pre COVID, during hospitalization, and post COVID timeframes for each patient in the patient summary node.

Data Cleaning:

- Filtered out unreasonable dates of birth that were prior to 01-01-1902 and after 01-01-2023. Controlled by min_reasonable_dob and dynamic function max_reasonable_dob which is the current date plus the max date shift for any data partner in the COHORT node.
- Corrected age_at_covid values filled with 0 to be null to avoid confusion with patients who have a valid birth date but are less than 1 year old.
- “race_ethnicity” variable has been updated to include the minimum 5 categories defined by the NIH and fix a bug where more specific race_concept_names were being improperly categorized as “Unknown”.

May 3, 2022:

Modifications:

- Condition_era input table was replaced with condition_occurrence table due to discrepancies between the patients found in the two tables as well as to provide full representation of facts in the visit level table.

Additions:

- Flags for PCR_AG_pos, PCR_AG_neg, Antibody_pos, Antibody_neg, and is_first_reinfection were added to the visit level fact table.
- A had_at_least_one_reinfection_post_covid_indicator was added to the patient level fact table. Defaulted to flag reinfection when there is a positive PCR or AG COVID-19 lab test > 60 days after the COVID patient’s index date but can be modified by input parameter reinfection_wait_time.

May 5, 2022:

Bug Fixes:

- during_macrovisit_hospitalization column in the visit-level dataset updated to flag dates that fall between macrovisit_start_date and macrovisit_end_date for any of the patient's macrovisits even if the date does not have a row in the microvisits_to_macrovisits table

May 9, 2022:

Additions:

- New Severity_Type flag created in the Patient Summary Table. Please note: this is NOT identical to the logic used by the Cohort paper team when they identified critical visits and assigned a severity score accordingly. This is similar in spirit with the following distinctions:
- The score is calculated for all confirmed positive, not just PCR/AG confirmed but also U07.1 diagnosed only.
- The only ED and hospital visits that count towards the score are those at the time of first infection.
- The Logic Liaison definitions of COVID-19 associated ED and hospital visits apply CDC developed logic to flag these visits.
- The value labels have been expanded to more explicitly reflect the TEMPORAL rather than CAUSAL nature of the (old and new) severity scores. For example, in this observational dataset, being hospitalized at the time of COVID-19 infection DOES NOT mean that COVID-19 was the reason for admission. Reference the Data Dictionary above to see the new logic.

May 18, 2022:

Modifications:

- Updated joins in the cohort_all_facts node to automatically join on any shared columns between the two datasets being joined. This will allow researchers to utilize input fusion sheet without need for manually accounting for shared columns from the _of_interest nodes when bringing all facts together for the final fact tables.

June 17th, 2022:

Modifications:

- Retain PCR_AG_Pos and PCR_AG_Neg flags in post covid timeframe, Antibody_Pos and Antibody_Neg flags in pre and post covid timeframes from cohort_all_facts node when collapsing to COVID_Patient_Summary_Table.
- Added PREGNANCY, OTHERIMMUNOCOMPROMISED, and SOLIDORGANORBLOODSTEMCELLTRANSPLANT comorbidity condition indicators for pre and post covid timeframes via fusion sheet input.
- Added PNEUMONIADUETOCOVID and MISC condition indicators in post covid timeframe via fusion sheet input.
- Updated select statements in COVID_Patient_Summary_Table to automatically join on any shared columns between the fusion sheet indicator_prefix column and cohort_all_facts to prevent errors when zero patients are flagged with a particular concept set in the fusion sheet.

July 15th, 2022:

Additions:

- Created flag for vaccination event in visit-level table and created counts of vaccinations pre and post covid index, using the output of the Vaccine Fact template
- Flag patients who were who were expired at the time of discharge or discharged to hospice from the visits domain as deceased

Modifications:

- Switched from using January 1st to July 1st as placeholder month and day of birth when there are 0s or nulls in the OMOP person table to avoid biasing towards older age.
- Separated the fusion input dataframe into two, one that can be completely customized and one that the required input rows for the template to run and create the proper derived variables in the patient summary node (LL_IMV and LL_ECMO)
- Re-introduced all possible visit dates from the visit_occurrence table for each patient (partial reversal of April 21, 2022 change where only retaining visit dates where fact being looked for is found)

August 4th, 2022:

Additions:

- Created flag death_within_specified_window_post_covid to reflect deaths that occurred within n days of the COVID index. This window is controlled by the reinfection_wait_time parameter in the cohort_all_facts_table node. The default is 60 days.

Modifications:

- Severity Type changed from Death_after_COVID_index to Death_within_n_days_after_COVID_index as a result of replacing COVID_patient_death =1 with death_within_specified_window_post_covid =1 as the qualification for this severity level.

September 7th, 2022:

- "race_ethnicity" variable has been updated to include new values of race_concept_name from latest data release on 9/1/22, including American Indian or Alaska Native Non-Hispanic

October 28th, 2022:

- Added PULMONARYEMBOLISM comorbidity condition indicator for pre and post covid timeframes via fusion sheet input
- Added B94_8 condition indicator in post covid timeframe via fusion sheet input

December 7th, 2022:

Additions:

- Added "race" variable using only race_concept_name from the OMOP person table. Created for researchers who do not necessarily want to use the combined "race_ethnicity" variable from the template.

Modifications:

- Updated column names to more accurately reflect the data that is represented in those fields. See Data Dictionary for more details.
  - "gender_concept_name" (now named "sex") OHDSI definitions and explanations are linked in the data dictionary entry.
  - "visit_date" (now named "date")

        - _before_covid_indicator (now named "_before_or_day_of_covid_indicator")
    - Updated observation period calculation to include index date as max for pre covid timeframe and min for post covid timeframe.
    - County level SDoH variables removed from L3 version of template due to new release of SDoH template that provides ZCTA level social determinants of health along with a select few county level variables not covered by the ZCTA level columns.

January 4th, 2023:

- Switched to using SUBSTANCE USE DISORDER concept set with appropriate renaming of indicator from prior concept set of SUBSTANCE ABUSE via fusion sheet input